

Plan van aanpak Stage - Big Data



Inhoudstafel

Inleiding	3
Stagebedrijf	4
Voorstelling stagebedrijf	4
Relevante aspecten voor de stage / Situering stage binnen de organisatie	4
Stageopdracht	5
Aanleiding en achtergrond	5
Doelstelling	5
Business case	6
Planning	7
Stappen/fasen van uitvoering + timing	7
Informatieverzameling en rapportering	8
Besluit	9

Inleiding

Bouw een Big Data Analytics platform, End-to-End solution, in de Cloud. Dat is de grote titel van de opdracht waar ik me als stagiair bij Ordina aan gewaagd heb.

Dit document bevat de volledige scope van deze opdracht in de vorm van een plan van aanpak. Uitleg over het stagebedrijf, de stageopdracht, de planning van de ingeplande topics die behandeld moeten worden en tenslotte informatieverzameling en rapportering van alles wat met de opdracht te maken heeft is te vinden in dit document.

Dit plan van aanpak beschrijft duidelijk belangrijke informatie over de verschillende fasen en taken die ik als stagiair bij Ordina zal ondergaan gedurende de komende dertien weken.

Stagebedrijf

Voorstelling stagebedrijf

Ik ga mijn stage doen bij Ordina. Ordina is een onafhankelijke IT-dienstverlener en een van de grootste consultancy-bedrijven in de Benelux. Hun motto is "Ahead of change". Zij willen hun partners helpen om digitaal te transformeren en de verandering voor te blijven. Ze doen dat door technologie, business-uitdagingen en mensen met elkaar te verbinden. Ze helpen vragende bedrijven hun snelheid te vergroten, slimme IT-toepassingen te ontwikkelen, nieuwe digitale diensten te lanceren en zorgen dat mensen al deze kunnen omarmen.

Het bedrijf bevindt zich op meerdere locaties binnen de Benelux met verschillende kantoren in Nederland (hoofdzetel), Vlaanderen, Wallonië en Luxemburg. Ik ben voor mijn stage bij praktisch elk van deze kantoren welkom, maar voornamelijk zal ik mij naar het kantoor van Mechelen begeven.

Om hun vele kantoren in de Benelux te bemannen zijn er momenteel maar liefst al meer dan 2500 werknemers onder contract bij Ordina met bijna duizend in Vlaanderen alleen en ze willen dat getal zo snel mogelijk overschrijden.

Deze enorme aantallen werknemers zijn natuurlijk opgedeeld in verschillende afdelingen. Deze worden omschreven als "Units" binnen Ordina. Zo is er bijvoorbeeld DataDriven, Robotics, Jworks, Business Platform Services, cybersecurity en nog veel meer. Deze Units hebben elk een hoekje toegewezen gekregen binnen de kantoorgebouwen, maar je bent absoluut niet gebonden om altijd bij jouw Unit te gaan zitten. Je mag namelijk gaan zitten waar je maar wil in de kantoorgebouwen. Dus ook bij andere Units zagezegd. Niemand heeft een vaste zitplaats. Vaak overlappen de verschillende Units met elkaar voor projecten waardoor je ongetwijfeld ook met andere Units nauw zal samenwerken in vele situaties.

Om af te sluiten, ze investeren nonstop in een 'great place to work' door fun, kennisdeling en samenwerken in teams. Daarom onderscheiden ze zich duurzaam bij de klanten, maar ook op de arbeidsmarkt.

Relevante aspecten voor de stage / Situering stage binnen de organisatie

Mijn stageopdracht situeert zich binnen de unit van DataDriven. Alles wat met Big Data te maken heeft, Data Engineering, Data Science, Data Visualization... wordt aan gesleuteld binnen deze Unit. Zo een component binnen DataDriven wordt binnen Ordina Practice genoemd. Mijn stage zal ik dus voor het grootste gedeelte binnen DataDriven doorbrengen waar ook mijn stagementor Aron Geerts, Big Data / Cloud Engineer, zich onder bevindt. DataDriven is te vinden op de eerste verdieping in het kantoorgebouw 3B in Mechelen.

Stageopdracht

Aanleiding en achtergrond

De stageopdracht is een volledig fictief Big Data project dat is uitgedacht door mijn stagementor Aron hemzelf. Deze opdracht is zo uitgedacht dat de stagiair alle hoofdonderdelen van Big Data een keer onder de loep kan nemen en kan samenvoegen tot een bepaald eindresultaat. Deze hoofdonderdelen zijn onder meer Data Ingestion, Data Engineering, Data Visualization en Data Science.

Wat is de aanleiding van de opdracht?

De fictieve luchthaven Vision Airport heeft een enorme hoeveelheid aan data dat moet geraadpleegd, verwerkt, gevisualiseerd en opgeslagen worden. Dit gaat bijvoorbeeld over alle data dat te maken heeft vliegtuigmaatschappijen, verschillende gates op de luchthaven, over de vliegtuigen en de vluchten zelf, het aantal personen per vlucht, weerinformatie, datums en nog veel meer. Omdat het echt wel van cruciaal belang is dat er zorgvuldig en op een veilige manier met al deze data wordt omgegaan moet er dus een centrale plaats voorzien worden waar al deze data veilig en betrouwbaar kan naartoe gestuurd, bewerkt en opgeslagen worden. Een end-to-end solution zal dus gevormd moeten worden

Deze stageopdracht zal volledig op individuele basis verlopen. Ik werk dus niet samen met nog andere stagelopers of werknemers aan deze opdracht. Natuurlijk kan en mag ik wel vragen stellen aan mijn stagementor Aron of andere werknemers binnen de DataDriven Unit indien er iets onduidelijk is of als ik problemen ondervind tijdens de technische uitwerking van het project.

Doelstelling

Er moeten dus heel wat technische zaken gerealiseerd worden om de bovenstaande situatie tot een mooi en effectief eindresultaat te vormen. Er zijn vier componenten die het eindresultaat moeten vormen:

I. Data inladen uit verschillende bronnen (files, API's...)

Alle benodigde data omtrent de luchthaven moeten we uit verschillende bronnen gaan halen en bestuderen. Deze data zal vooral gehaald worden uit allerlei soorten files waar luchthaveninformatie in opgeslagen en verzameld is. Maar indien mogelijk zal er ook data worden verzameld met behulp van API's. Om bijvoorbeeld live weersomstandigheden te verzamelen kunnen we gebruikmaken van gratis bestaande weer API's.

Onder deze component vallen de volgende topics uit de planning: opstart + data exploratie, API data extractie (later).

II. Ontwikkelen van een Data Lake

We willen graag die enorme hoeveelheid data die voortkomt uit de files en eventueel API's ergens op een centrale plek opslaan. Een toegankelijke maar betrouwbare plek. Daarom zullen we een Data Lake moeten ontwikkelen in AWS. Deze soort van Data opslag heeft ook geen vast schema waaraan alle files en data die binnenkomt zich moet houden. We moeten ons dus geen zorgen maken over ongestructureerde data die binnenkomt. Deze ongestructureerde data zal later dan omgevormd maken naar data die ordelijk zinvol en gestructureerd is zodat gewenste data simpel opgevraagd kan worden door wie dat wil.

Onder deze component vallen de volgende topics uit de planning: Landing + Raw layer, Modelleren data + Cleansed zone.

III. Rapportering

Wanneer we bij deze component belanden hebben we alle data al opgeslagen en herschreven in het Data Lake maar er nog niet echt iets zinvols mee gedaan. We gaan dus alle data die belangrijk is voor de luchthaven op een bepaalde manier visualiseren. Met behulp van Dashboards en Interfaces kunnen

we dan (waarschijnlijk in de AWS omgeving) zinnvolle en gewenste real-time data op een aantrekkelijke en ordelijke manier visualiseren.

Onder deze component vallen de volgende topics uit de planning: Rapportage + dashboard.

IV. Data Science toepassen

Nu we bijna alle componenten van Big Data hebben toegepast binnen de opdracht mist er nog één belangrijke, Data Science! We willen aan de hand van de real-time data dat we allemaal al hebben verzameld over de luchthaven voorspellingen gaan maken. Hier zal een stukje A.I. en Machine Learning bij komen kijken. Aan de hand van historische data i.v.m. vluchten en de invloed van het weer op deze vluchten zullen we dus proberen te voorspellen hoeveel vertraging er zal zijn in bepaalde weersituaties. Heel interessant om toe te voegen aan het geheel! Zo kan de luchthaven anticiperen op bepaalde situaties die voorkomen.

Onder deze component vallen de volgende topics uit de planning: Data Science toepassing, API data extractie (weer-API).

Wanneer al deze stappen succesvol tot een eindresultaat zijn uitgewerkt en een mooi werkend geheel vormen is er nog altijd ruimte voor uitbreidingen en extra's die het geheel efficiënter of aantrekkelijker maken indien er nog wat dagen/weken resteren.

De technische aspecten van de componenten, de bijhorende geplande topics, hoe deze zijn onderzocht en toegepast zullen in de projectdocumentatie, die zal gemaakt worden gedurende de hele realisatiefase, uitgebreid beschreven worden.

Business case

Waarom is deze stageopdracht interessant voor Ordina zelf?

Ordina wilt vooral de stagiair zoveel mogelijk in contact te laten komen met de belangrijkste technische aspecten van Big Data en zo de stagiair veel laten bijleren. Wanneer deze stageopdracht op een goede manier tot een goed einde wordt gebracht is het voor Ordina interessant om te kijken of deze stagiair bekwaam genoeg is en past bij het team van Ordina zelf. Deze stage is dus de beste sollicitatie om te kijken of ik een waardige toevoeging ben aan het team en zo officieel tot Ordinees wordt benoemd.

De stageopdracht is een vergelijkbare simulatie met een echt project waarvoor Ordina zou ingehuurd worden. Als ik deze opdracht tot een mooi eindresultaat kan vormen en ook een leuke fit kan zijn voor het team tijdens de stage, is de kans dat ik echt als werknemer bij Ordina een waardige toevoeging zou zijn reëel.

Planning

Stappen/fasen van uitvoering + timing

In onderstaande tabel staan alle topics die gepland zijn gedurende de stageperiode met hun geschatte duur in weken opgesomd.

Tijdsbesteding stage	
Topic	Geschatte duur (in weken)
Opstart + data exploratie	2
Landing + Raw layer	1
Modelleren data + Cleansed zone	3
Rapportage + dashboard	1
Data Science toepassing	2
API data extractie	1
Samenbrengen resultaten	1
(Redacted)	(Redacted)
(Redacted)	(Redacted)
Totaal	11

Informatieverzameling en rapportering

Om de inleiding te herhalen, voor een goede uitvoering van de opdracht is het belangrijk dat ik de stagebegeleider (Docent Dirk Mervis) en stagementor (Aron) frequent op de hoogte hou van het verloop van de stage op technisch en psychologisch vlak.

Voor deze opvolging van feiten zullen de volgende acties ondernomen worden:

✓ Plan van aanpak

Het plan van aanpak zorgt voor een duidelijk overzicht van de volledige scope van de stageopdracht. Dit maakt het zowel voor anderen als mezelf makkelijk om belangrijke info omtrent mijn stageopdracht terug te vinden op een centrale plaats (dit document). Bedenkingen over het stagebedrijf, de opdracht zelf, de planning van de componenten waaruit de opdracht is opgesteld en informatieverzameling en rapportering kan gemakkelijk teruggevonden worden voor wie dat wenst.

✓ Persoonlijk logboek

In dit logboek verschijnen dagelijks updates van uitgevoerde werkzaamheden en ervaringen. Zo is het voor anderen en mezelf simpel om terug te kijken op verschillende periodes gedurende de stageperiode indien nodig.

✓ Projectdocumentatie

Natuurlijk is het belangrijk dat alle technische bevindingen en ervaringen doorheen de volledige realisatiefase van de opdracht ordelijk en gestructureerd gedocumenteerd worden. Dit gaat dan vooral over de voor mij geplande topics die ik moet uitwerken tijdens de stageperiode. Zo zal duidelijk worden hoe ik welk topic heb onderzocht en op welke manier ik deze heb uitgewerkt in de opdracht.

✓ Wekelijkse meeting met stagementor

Om Aron (stagementor) goed op de hoogte te houden van de huidige wekelijkse situatie waarin ik me bevind houden we minimum één meeting per week om alles op een rijtje te zetten. Zo is hij ook frequent up-to-date waar ik met de opdracht al sta en wat er nog moet gebeuren. Als er echt dringende vragen zouden zijn van zowel Aron als mezelf kunnen we elkaar elk moment bereiken via MS Teams of via e-mail.

Aron en ik hebben ook een gezamenlijke Trello-workspace en GitHub repo zodat hij elke dag kan volgen waarmee ik bezig ben (Trello) en de geschreven code kan terugvinden (GitHub).

✓ Terugkommomenten op Thomas More

Er zijn ook twee terugkommomenten voorzien op Thomas More in Geel. We komen dan samen met de stagebegeleider (Docent Dirk Mervis) en de andere studenten van Thomas More die stagelopen bij Ordina om te overlopen waar we elk individueel aan bezig zijn en hoe we de stage ervaren.

Besluit

Dit plan van aanpak zou een duidelijke scope moeten geven van de belangrijke aspecten van de volledige stageopdracht. Uitleg over het stagebedrijf, de stageopdracht, de planning van de hele periode en de bijhorende geplande topics en tenslotte informatieverzameling en rapportering beschrijft duidelijk wat er in de komende maanden verwacht wordt van mij als stagiair. Het is een enorm interessante opdracht en ik ga zoveel bijleren dankzij de toch wel vele uitdagingen die ik onderweg zal tegenkomen. Zeker als Cloud & Cybersecurity student gaat het een leuke en interessante verrijking zijn van ervaringen in een ander gebied binnen IT!

Ik kijk er enorm naar uit om mijn best te doen voor deze stageopdracht, voor Aron en voor Ordina!

#DataDriven